R E P O R T R E S U M E S

ED 002 972 24

PROBLEM-SOLVING PROFICIENCY AMONG ELEMENTARY SCHOOL TEACHERS, III. TEACHERS OF READING GRADES 2-5.

BY- WADE, EUGENE W.
INDIANA UNIV., BLOOMINGTON, INST. OF EDUC. RES.

REPORT NUMBER CRP-419-3 PUB DATE JAN 61
EDRS PRICE MF-$0.09 HC-$1.84 46P.

DESCRIPTORS- *TEST VALIDITY, *SKILL DEVELOPMENT, TEST CONSTRUCTION, READING COMPREHENSION, *READING INSTRUCTION, TEACHING QUALITY, EFFECTIVE TEACHING, *TEACHER EVALUATION, ELEMENTARY SCHOOL TEACHERS, INTELLIGENCE, *PROBLEM SOLVING, BLOOMINGTON, INDIANA

A RELIABLE NORMATIVE TEST COVERING A SAMPLE OF TEACHING SKILLS IN READING INSTRUCTION FOR GRADES 2-5 WAS CONSTRUCTED AND VALIDATED. THE HYPOTHESIS WAS MADE THAT TEACHERS BECOME MORE SKILLFUL IN SOLVING TEACHING PROBLEMS AS THEY GAIN EXPERIENCE. IT WAS SUGGESTED THAT A VALID TEST OF TEACHER SKILLS IN READING SHOULD THEREFORE REFLECT THIS PROCESS. SAMPLES OF TEACHERS, STUDENT TEACHERS, AND EDUCATION UNDERGRADUATES WERE TESTED AND COMPARED, CROSS SECTIONALLY, ON EACH OF NINE SUBTESTS AND ON A TOTAL TEST SCORE. ALL COMPARISONS REVEALED HIGHER PROFICIENCY AMONG THE MORE EXPERIENCED ON THE TOTAL TEST SCORES, ALTHOUGH SOME DISCREPANCY WAS FOUND IN THE SUBTESTING. SECONDARY CORRELATIONS WERE MADE BETWEEN THE READING-PROBLEMS TEST AND SELECTED INDEPENDENT VARIABLES, PARTICULARLY INTELLIGENCE AND READING COMPREHENSION. RESULTS ALLOWED PRELIMINARY VALIDATION OF THE TEST INSTRUMENT. THIS STUDY WAS DEVELOPED AS THE THIRD OF SEVERAL REPORTS ON COOPERATIVE RESEARCH PROJECT 419. (JH)

ED002972

REFERENCE
COPY

COOPERATIVE
RESEARCH
PROJECT

No. 419-III

419

MONOGRAPH
of the
INSTITUTE OF EDUCATIONAL RESERACH
at
INDIANA UNIVERSITY

Problem Solving Proficiency Among Elementary School Teachers

III. Teachers of Reading Grades 2-5

by
Eugene W. Wade

JANUARY 1961

PROBLEM SOLVING PROFICIENCY AMONG ELEMENTARY SCHOOL TEACHERS

III. TEACHERS OF READING

GRADES 2-5

by

Eugene W. Wade

The research reported herein was supported through the Cooperative Research Program of the Office of Education, U.S. Department of Health, Education, and Welfare

Published by the
INSTITUTE OF EDUCATIONAL RESEARCH
School of Education
Indiana University

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Price, $1.00, postpaid
For sale by the Institute of Educational Research
School of Education
Bloomington, Indiana

TABLE OF CONTENTS

	Page
INTRODUCTION	1
PROCEDURES	3
FINDINGS	13
Validation	13
SUMMARY	35

LIST OF TABLES

Table		Page
1.	Sentence Construction In Each Excerpt of Problem 4 . . .	6
2.	Word Difficulty Of Each Excerpt In Problem 4	6
3.	Southern Indiana Teachers, 1-6 Years Experience, Vs. Indiana University Undergraduates	13
4.	Total Chicago Teachers Vs. Indiana University Undergraduates	15
5.	Indiana University Student Teachers Vs. Indiana University Undergraduates	16
6.	Combined Teacher Group Vs. Indiana University Student Teachers	17
7.	Performance of Chicago Teachers Who Have Taught 1-10 Years, 11-20 Years, 21-30 Years, 31-43 Years	19
8.	Comparison of Chicago Teachers Who Have Taught 1-15 Years and 16-30 Years	20
9.	Performance of Southern Indiana Teachers Who Have Taught 1-10 Years, 11-20 Years, 21-30 Years, 31-40 Years .	21
10.	Means, Standard Deviations, and Split-Half Reliability of Forms AA and AB of the Reading-Problems Test . . .	22
11.	Chi Square Median Test Values for Each Item Using The Combined Teacher Group	23
12.	Correlation Coefficients of Each Subtest With the Total Score for All Samples	26
13.	Total Score Variances for All Samples	27
14.	Correlation Matrix of Combined Teacher Sample (N = 176).	28

Table		Page
15.	Eigenvalues for Each Subtest and Total	29
16.	Subtest Factor Loadings, Factor Variances, and Communalities From the Varimax Rotation	30
17.	Description of the Solution Process Involved in the Sub-tests Which Load on Each Factor	31
18.	Correlation Matrix of the Total Score of the Reading-Problems Test, ACE, Co-op C-2, MTAI, and Grade-Point Average on 45 Indiana University Student Teachers . .	33
19.	Summary of Results of the Validation, Reliability, and Internal Consistency Studies For the Total Score and Subtests	36

INTRODUCTION

The purpose of the present study was to explore the validity and reliability of nine problems in the teaching of reading in grades two through five. That the purpose was to explore rather than establish the validity of these problems rested largely on the considerations previously set forth by Turner and Fattu (10, 11), namely, that the validation of constructed problems by which to assess teaching skills requires the use of many criteria, rather than a single criterion in which full confidence is invested.

For the initial exploration of the validity of problems in teaching reading, the author chose two criteria. The first was the differentiation of elementary teachers with respect to degrees of training and experience by means of constructed problems. The second was sensibility or reasonableness of the relationships between problem performance and other performance measures, including other tests.

With respect to the first of the above criteria, the problem was to show that there are differences in performance between persons with no training and no experience in teaching reading (e.g. sophomores enrolled in an elementary education curriculum) and persons with some training and some experience in teaching reading (e.g. graduating seniors in elementary education), and then between these two groups and experienced teachers. Unless experienced teachers, who presumably

practice the skills involved in the problems constructed, outperform persons who have had little occasion to practice them, one may place little confidence in the problems as measures of a specialized skill in teaching.

With respect to the second of the two criteria mentioned, the problem was largely one of having to obtain reasonable or explainable relationships between problem solving performance in teaching reading and other types of performance. For instance, a test of skills in teaching reading could reasonably be expected to hold moderate positive relationships to reading comprehension. Moreover, insofar as it is expected that skills in teaching reading are learned, it is reasonable to suppose that there should be at least a small positive relationship between problem performance and learning ability as measured by a test of intelligence. Beyond these two variables, the exploration of the relationships in the second criterion is largely an empirical matter. In the present study, grade point average in undergraduate preparation and MTAI scores were examined on a largely empirical basis in relation to problem performance.

A matter somewhat tangential to the initial efforts to validate problems in teaching reading and yet relevant to gaining better understanding of such problems and to the ultimate interpretation and use of them is the factorial relationships between the problems. The question dealt with in this context was whether skill in teaching reading is homogeneous and is thus largely accounted for by a single factor or

whether it is heterogeneous and involves several separate factors.

PROCEDURES

Prior to construction of problems, several steps were taken. Among these were: (1) the objectives in reading stated by Greene and Gray (7) were examined; (2) classroom observations of teacher problem solving activities while teaching reading were informally made; (3) statements of skills required of reading teachers, such as those by Gray (6) and Gates (4) were studied.

The reading-problems test was designed to measure the skills believed to be necessary for teachers to help children select suitable reading books, to place children into homogeneous reading groups, to judge pupil improvement after classroom instruction, to diagnose and correct a child's phonic and syllabication errors, to organize a child's word perception errors into categories that will be meaningful for instruction, and to recognize the goals implicit in various reading workbook exercises.

Problem description. The test initially consisted of ten subtests or problems. Six of these were tape recordings of passages from various books and four were paper and pencil problems. During the preliminary testing of the problems, one of the six problems was deleted. A description of the nine remaining problems follows:

Subtest 1---designed to measure ability to consistently use the accepted criterion of difficulty in selecting the proper level reading book for different readers. The instructions to the teachers were:

> "First: You are the teacher of Paul, Pam, Richard, and Connie. You desire to select books for instruction in the classroom.
>
> Second: You will hear each child read aloud the three selections (beside his name) which he has just finished reading silently.
>
> Third: After each selection, decide if the material of that difficulty level is too easy, OK, or too hard for that child's instruction in the classroom."

The scoring scheme for this first problem was objective. The number of words miscalled was taken as the criterion of difficulty. A selection was at the child's reading level when he averaged one error per twenty running words. In the twelve readings only one met this criterion. The other eleven were either definitely above or below it.

Subtest 2---designed to measure ability to consistently recognize different levels of reading proficiency and thus assign children to homogeneous reading groups. The instructions were:

> "You will hear ten children read aloud. Your task will be to form three instructional reading groups. Each will read only once and all will read the same paragraph. Group them on the single criterion of oral reading accuracy."

In problem two, the number of seconds taken to read the selection plus the total number of errors determined the correct grouping.

Subtest 3---designed to measure ability to judge improvement or

no improvement in reading with the passage of time and learning experiences. The instructions were:

> "Listen to the following recorded oral readings of children taken before and after three months of classroom instruction in reading. Judge the growth that each has made in reading."

Each child actually read the selection only once -- this one selection was taped twice and represented as the "before" and "after" performance with "three months of classroom instruction" between.

Subtest 4---designed to measure ability to inspect several books and know how they rank in reading difficulty for children.* This skill was one which a teacher would use in helping children select appropriate level library books. The instructions were:

> "First: Rank all the following excerpts for difficulty. Number the easiest #1 and the most difficult #5."

In problem four, the correct ranking was determined by two independent methods. First, by the construction of the paragraphs as indicated in Tables 1 and 2.

---

*This problem was suggested and constructed by R. L. Turner.

TABLE 1. SENTENCE CONSTRUCTION IN EACH EXCERPT OF PROBLEM 4.

Paragraph	Number of Sentences	Range of Sentence Length	Mean Sentence Length	Median Sentence Length	Number of Prepositional Phrases
A	13	2 - 15	7.69	6.0	4
B	10	6 - 16	9.50	9.5	5
C	9	4 - 17	11.22	11.5	4
D	8	5 - 21	13.37	12.5	8
E	6	3 - 23	17.16	21.5	12

TABLE 2. WORD DIFFICULTY OF EACH EXCERPT IN PROBLEM 4.

Paragraph	Words in Thorndike's (15) First 1000	2nd 1000	3rd 1000	4th 1000	5th 1000	6th 1000	7th 1000	above 7th 1000
A	87	2	4	1	0	0	0	0
B	82	6	5	1	3	0	0	0
C	83	7	3	3	3	1	0	0
D	81	8	8	2	2	2	1	1
E	74	12	1	5	2	1	3	2

The second technique used to find the correct ranking was to have six children read the five paragraphs. Their total errors and reading time were tabulated. The difficulty order indicated by the

errors was the same as the order by construction.

Subtest 5---designed to measure the ability to predict which of several books would be appropriate for a particular child without his reading the book aloud. This skill was necessary in selecting reading books for children. The instructions were:

"First: John will read once. He will read only this paragraph.

> Tom and Ned live near a large city park. They often visit it with their playmates. In the park are many shady maple trees. There is a pleasant picnic ground on the hill, and the valley below has a pretty little pond. The girls always enjoy watching the boys while they sail their tiny boats in the water. Mother and Father enjoy picnics in the park. (5)

Second: Form an opinion of how well John reads the paragraph.

Third: Look at the excerpts in Problem 4 (the five that you just finished ranking).

Fourth: Make an educated guess about which of these five excerpts John would be able to read comfortably (averaging one error in twenty running words).

Fifth: Remember -- you have to make your guess on the basis of how well he reads the paragraph at the top of this page -- he will not read any of the excerpts.

Sixth: Write the letter of the excerpt which you predict John would be able to read comfortably beside his name below.

Seventh: Get ready to repeat the procedure with Mary Lou, Karen, Eddie, and Pam."

This problem was scored by having each of the children read the five excerpts. The correct answer was the excerpt which each

reads with an average of one error per twenty running words.

Subtest 7---designed to measure ability to give general rules of phonics to children as they made vowel, consonant, and combination errors. The instructions were:

> "First: Listen carefully as Sheila tries to sound out each of the eight words below.
>
> Second: When you hear an error, write beside that word the general rule which she should learn to prevent future mistakes of that sort."

Four common rules were required to correct Sheila's eight errors. These cover the cases of double vowels, silent e's, a single vowel word, and the blending of a consonant with r.

Subtest 8---designed to measure ability to give general rules of syllabication to children who made errors in dividing words so that future mistakes of this sort will be avoided. The instructions were:

> "First: The diagonal lines in the words below show where Mary broke the words into syllables.
>
> Second: Write beside those words in which she made an error the rule that Mary needs to know to prevent future mistakes of that sort.
>
> Third: When a word is correctly divided, write correct and state no rule."

The eleven words in which Mary made errors called for the four rules of syllabication which deal with compound words, prefixes and suffixes, single syllable words and the consonant which follows a long or short vowel.

Subtest 9---designed to measure ability to conceptualize or categorize a reader's perception errors on sight words as a preparatory step in the child's instruction. The instructions were:

> "The following list of 36 sight words was shown to Mike on flash cards with an exposure time of one second per word. He made 10 errors which his teacher wrote in (on the list below) beside the correct word. These 10 errors can be classified into different groups which will be meaningful for his future instructions.
>
> First: Look over the ten errors and decide what are the different kinds of errors.
>
> Second: Put each kind into a different space in the table below. Instead of writing the words in the spaces, use the number of the word.
>
> Third: Make no fewer than two groups and no more than eight.
>
> Fourth: Label each of the kinds of errors."

The errors here fall into five categories: initial, middle, and end errors; reversals; and complete word substitution. Partial credit was given for certain mixtures under more general concepts.

Subtest 10---designed to measure ability to assign the correct exercise after the child's weaknesses were identified. This required that the teacher know what skills were developed by what exercises. An indirect way of measuring this teacher skill was to test her ability to recognize similarity of intent between exercises designed to build the same or an associated skill. The instructions were:

> "First: Read over the six sample exercises printed below, each of which is designed to build a particular reading skill.

Second: On the opposite page are nine exercises of the same type; all of which you are to match with the six below.

Third: Place beside each sample below, the number of the exercise or exercises from the opposite page which build the same or an associated reading skill.

Note: You are to use all of the 9 exercises on the opposite page but each may be used only once."

Problem ten was constructed by taking samples of two or three like exercises from Russell and Karp's Reading Aids Through the Grades (8), and Bond and Tinker's Reading Difficulties (1), Their Diagnosis and Correction (2). These samples were modified to fit the test. A response was marked correct when like exercises of a sample were matched.

Sampling. Four samples were drawn as follows:

1. One hundred and seven teachers of all ages and experience who taught in grades two through five in a southern Indiana school system. A sub-sample was taken from the sample. It consisted of 45 teachers who had from one to six years of teaching experience. (Because of selection factors, teachers with more than six years experience in the southern Indiana sample were not representative of the teacher population. See Tables 7 and 9.)

2. Sixty-nine Catholic sisters. These nuns taught in grades two through five in Chicago parochial schools. The population of nuns was sampled because nuns typically do not leave their profession, hence there is little opportunity for differential selection to operate as years

of teaching experience increases. (The "combined teacher sample", n = 176, was the combination of these first two samples.)

3. Ninety-seven students enrolled in a sophomore course in child development in a midwestern university during the school year 1959-60. These undergraduates had neither teaching experience nor a methods course in teaching reading.

4. Eighty-nine seniors in elementary education who had completed 16 weeks of student teaching at a midwestern university.

Determination of reliability. The split-half reliability technique with the Spearman-Brown correction was used. The test was divided into two equivalent forms of equal difficulty and discriminatory power on the basis of an item analysis of the responses of the southern Indiana, Chicago, and undergraduate samples.

Item Analysis. The item analysis was made in two ways by the chi square median test. Each item's discriminatory power was tested against: (a) the total score, and (b) the subtest score. The combined teacher group (n = 176) was used.

In the first case (item vs. total score), it was reasoned that the worth of an item is related to the degree that it will discriminate those subjects whose total score is above the median from those whose score is below the median.

In the second case, an item's worth was in relation to its power to discriminate between those who score above and below the median of the subtest of which that item is a part. It was judged that an item which did not contribute to its own subtest probably should be discarded.

Subtest analysis. The subtest analysis was done by product moment correlations. Each subtest was correlated with the total score. To be sure that the relationships held in different samples, separate correlations were computed for the southern Indiana, Chicago, and combined teacher groups; and the student teacher and undergraduate groups.

Factor analysis. In order to learn as much about the test as possible, a factor analysis and a four-fold varimax rotation were computed. This factor analysis, like the item analysis, was done on the combined teacher group (n = 176).

Relationship to other tests. To yield information about the relationship between the reading problems test and other types of performance of possible relevance to teaching reading, correlations were computed between the total score and the ACE psychological examination (9), the Coop C_2 test of reading comprehension (3), the Minnesota Teacher Attitude Inventory (2), and college grade-point-average. These relationships were based on a sample of 45 student teachers. This was the number, from the total of 89, for whom all data on all variables was available.

FINDINGS

Validation

Southern Indiana teachers compared to undergraduates. Southern Indiana teachers with one to six years experience were compared to Indiana University undergraduates. It was hypothesized that teachers would outscore undergraduates on each subtest as well as on the total score. The results of the comparison are given in Table 3.

TABLE 3. SOUTHERN INDIANA TEACHERS, 1-6 YEARS EXPERIENCE, VS. INDIANA UNIVERSITY UNDERGRADUATES.

Subtest	Means, Teachers	Means, Undergraduate	Difference	Critical Ratio
1	6.36	6.16	.20	.67
2	8.47	7.95	.52	2.40**
3	.42	.04	.38	1.34
4	3.62	3.01	.61	2.28*
5	1.96	2.22	-.26	.59
7	6.33	1.19	5.14	17.36***
8	5.60	2.40	3.20	9.14***
9	5.04	3.28	1.76	3.23***
10	6.87	4.70	2.17	9.48***
TOTAL	44.67	30.85	13.82	11.15***

*Significant at .05 level
**Significant at .02 level
***Significant beyond .01 level (two tailed test)

The hypothesis holds up well except for problems one, three, and five. Three interpretations are possible. First, these three problems measure skills which are already fully developed by the time one becomes a college undergraduate. Second, the three subtests represent skills which teachers should have developed but have not. Problem one, for example, measures the teacher's ability to select the proper difficulty level reading book. If the majority of teachers have the whole class reading from the same book, skill in selection cannot develop; thus, teachers as a group would not significantly outscore undergraduates. Third, the undergraduate sample which is being used for the comparison may be superior to the total undergraduate population of teacher training institutions. Therefore, differences which may in fact exist may not be apparent.

Chicago teachers (nuns) compared to undergraduates. For the purposes of cross validation, the total Chicago group (n = 69) was compared to the Indiana undergraduates (n = 97). The results are shown in Table 4.

With the exception of problem two, those subtests which separated southern Indiana teachers from undergraduates also separated Chicago teachers from the undergraduates. However, problems one and five showed significant differences in the opposite direction; that is, undergraduates significantly outscored Chicago teachers on subtests one and

five. Besides the possible explanation that the undergraduate sample was not representative, it is also possible that in the particular skills assessed by those two problems, teachers as a group became less skillful with experience. A teacher who year after year has handed the same reading book to every child in her class regardless of his reading abilities or reading level, could develop wide tolerance for what is suitable. If this was in fact true, the reversed differences between the teachers and the undergraduates become understandable and do not constitute evidence against the validity of subtests one and five.

TABLE 4. TOTAL CHICAGO TEACHERS VS. INDIANA UNIVERSITY UNDERGRADUATES.

Subtest	Means, Teachers	Means, Undergraduates	Difference	Critical Ratio
1	5.20	6.16	-.96	2.90**
2	8.29	7.95	.34	1.44
3	.15	.04	.10	.56
4	3.55	3.01	.54	2.04*
5	1.32	2.22	-.90	2.25*
7	4.96	1.19	3.77	12.20**
8	6.62	2.40	4.22	11.17**
9	6.58	3.28	3.30	6.61**
10	5.45	4.70	.75	2.31
TOTALS	42.13	30.85	11.28	7.35**

*Significant at .05 level
**Significant beyond .01 level (two tailed test)

Indiana University student teachers compared to undergraduates.

To further investigate the effect of experience on test performance, student teachers who had completed 12 weeks of teaching were compared to undergraduates. It was hypothesized that student teachers would score higher than undergraduates who have had no experience. The results are shown in Table 5.

TABLE 5. INDIANA UNIVERSITY STUDENT TEACHERS VS. INDIANA UNIVERSITY UNDERGRADUATES.

Subtest	Means, Student Teachers	Means, Undergraduates	Difference	Critical Ratio
1	5.99	6.16	-.17	.56
2	8.21	7.95	.26	1.20
3	.06	.04	.02	.09
4	3.19	3.01	.18	.74
5	1.69	2.22	-.53	1.31
7	3.61	1.19	2.42	7.70**
8	4.19	2.40	1.79	5.57**
9	5.17	3.28	1.89	4.92**
10	5.88	4.70	1.18	4.78**
TOTALS	37.78	30.85	6.93	5.21**

**Significant beyond .01 level (two tailed test)

The subtests which showed rapid growth with the experience of

student teaching were those measuring phonic and syllabication rules, concept formation on word perception errors, and the matching of reading exercises. None of the first five subtests showed significant differences.

Combined teacher group compared to student teachers. While it is reasonable to expect that student teachers would score higher than undergraduates, they in turn should be outscored by experienced teachers. The comparison was made between the combined teacher group (n = 176) and the Indiana University student teachers (n = 89). The results are shown in Table 6.

TABLE 6. COMBINED TEACHER GROUP VS. INDIANA UNIVERSITY STUDENT TEACHERS.

Subtest	Means, Teachers	Means, Student Teachers	Difference	Critical Ratio
1	5.70	5.99	-.29	1.29
2	8.48	8.21	.27	1.39
3	.11	.06	.05	.39
4	3.50	3.19	.31	1.41
5	1.55	1.69	-.14	.40
7	5.38	3.61	1.77	5.60**
8	5.84	4.19	1.65	5.92**
9	5.40	5.17	.23	1.30
10	5.99	5.88	.11	.47
TOTALS	41.90	37.78	4.12	3.30**

**Significant beyond .01 level (two tailed test)

Only the total score and the phonics and the syllabication subtests showed significant differences. If the teacher and practice teacher groups were representative samples, the difference of only 4.12 between the means of their total scores (as compared to the difference of 6.92 between the undergraduate and student teacher means) suggested that teachers did not change much in their ability to solve these reading problems after they get on the job.

Test performance as a function of the years of teaching experience--Chicago teachers. If one predicts gain in problem solving skill with experience, it follows that within any teacher group those teachers with the greater number of years in the classroom should score higher than the less experienced teachers.

There are, however, factors which may influence the results of any cross-sectional study with teachers. Validation and normative studies on the MTAI (Form X-164) have shown that teachers over age 40 scored significantly lower than those under 40. Perhaps this change may be reflected in performance on the reading-problems test. A second factor which could influence a cross-sectional study of gain in test performance with experience is the change in instructional techniques emphasized during the past 40 years. For example, teachers trained during the 1930's when "whole word" techniques were advocated may not have the teacher skills in phonics and syllabication. The third, and most

significant factor which could affect a cross-sectional study, is the migration and loss from the profession of public school teachers. Turner and Fattu (10) have hypothesized that the best solvers of teaching problems tend to be found teaching in school systems where they enjoy a broad professional autonomy. If this hypothesis is true, the choice of a school system in which to demonstrate test gain with experience is of extreme importance. One must select a system from which teachers do not migrate. Therefore, the prediction of increased test performance as a function of experience within a teacher sample was tested using the Chicago teachers since Catholic nuns usually do not migrate or quit. The factors of changing attitudes and differences in preparatory training were allowed for by testing the prediction only up to that experience level at which performance deteriorates.

Table 7 shows the results of the gain-with-experience hypothesis with the years of teaching arbitrarily divided into ten year segments.

TABLE 7. PERFORMANCE OF CHICAGO TEACHERS WHO HAVE TAUGHT 1-10 YEARS, 11-20 YEARS, 21-30 YEARS, 31-43 YEARS.

Years experience	Number	Mean
1-10	27	41.74
11-20	19	45.53
21-30	12	45.83
31-43	11	34.00

The group trend here supports the prediction (except beyond 30 years experience) that higher test performance is associated with greater teaching experience.

If those teachers with more than 30 years of experience are excluded and the experience range 1-30 is divided into two groups 1-15 and 16-30, the contrast becomes greater as shown in Table 8.

TABLE 8. COMPARISON OF CHICAGO TEACHERS WHO HAVE TAUGHT 1-15 YEARS AND 16-30 YEARS.

Years Experience	N	σ^2	Mean	t	p
1-15	38	95.79	42.05	2.17	.05
16-30	20	61.86	47.20		

The severe drop in performance after 30 years of experience may be the result of changed attitudes or different training or it may only be sampling error. Because of the small number of cases in this section of the study a replication would be desirable.

Test performance as a function of the years of teaching experience--southern Indiana teachers. For comparison, the southern Indiana public school teacher sample was similarly divided into 10 year segments. The hypothesis of test performance improvement as a function of years of teaching experience was not made for this sample. Results are shown in Table 9.

TABLE 9. PERFORMANCE OF SOUTHERN INDIANA TEACHERS WHO HAVE TAUGHT 1-10 YEARS, 11-20 YEARS, 21-30 YEARS, 31-40 YEARS.

Years Experience	Number	Mean Total Score
1-10	58	44.33
11-20	22	41.05
21-30	17	36.47
31-40	10	39.30

The trend is toward decreasing performance with increasing experience. The combined effect of selective teacher loss and changes in attitudes and preparation are advanced as possible explanations. However, it is impossible to indicate the relative effects of these factors within the present sample.

Determination of reliability. Each of the nine subtests was divided into halves. The object was to make the halves of each problem as nearly equal as possible in discriminatory power, difficulty, and content. For example, in subtest eight--syllabication rules--each half contains a prefix word, a single syllable word, a consonant after vowel word, and a suffix word. The number of errors made on each half by the Chicago, southern Indiana, and undergraduate subjects was about equal. The power of each half to separate those who scored above and below the median score of the total test was also about equal.

The subtest halves were then assigned to alternate forms of the test. One form was designated FORM AA; the other FORM AB. These forms were correlated and the Spearman-Brown correction was made. Results may be seen in Table 10.

TABLE 10. MEANS, STANDARD DEVIATIONS, AND SPLIT-HALF RELIABILITY OF FORMS AA AND AB OF THE READING-PROBLEMS TEST.

Form	Mean	Standard Deviation	$r_{aa, ab}$	r corrected
AA	21.500	5.024		
AB	20.765	5.144	.759	.868

The reliability of either subtest is probably not sufficiently high for it to be used for individual measurement. However, the halves could be used as pre-test and post-test when only mean group gains are to be considered.

Item and Subtest Analyses. An item that will not distinguish significantly between subjects who score above and below the median of the subtest of which the item is a part is usually too poor a discriminator to remain in the test. There are two exceptions to this: (a) when the item is necessary to the subtest because it serves as a referent (this is true of the items in problem two where the teachers must use all of the pupils to establish the criteria for grouping); and (b) in those few cases

where an item fails to discriminate in the subtest but separates those who score above and below the median on total score.

Table 11 shows the chi square values for the items against subtest score and against total score.

TABLE 11. CHI SQUARE MEDIAN TEST VALUES FOR EACH ITEM USING THE COMBINED TEACHER GROUP.

Subtest Number	Test Item	Degrees Freedom	Chi-Square to Subtest Median	Chi-Square to Total Score Median
1	1(Paul)	1	1.64	3.71*
	2 "	1	.51	1.00
	3 "	1	2.58	4.14**
	4(Pam)	1	37.98***	.57
	5 "	1	38.02***	.00
	6 "	1	28.83***	.26
	7(Richard)	2	11.07***	10.53***
	8 "	2	9.37***	1.86
	9 "	1	36.20***	4.57**
	10(Connie)	1	.00	1.82
	11 "	1	27.44***	.92
	12 "	2	25.89***	.32
2	1(Ray)	1	12.58***	.35
	2(Stevie)	1	38.23***	1.37
	3(Patty)	1	4.00**	4.09**
	4(Danny)	1	12.88***	1.87
	5(Donna)	1	5.93**	4.65**
	6(Irvin)	1	12.71***	1.13
	7(Buddy)	1	16.34***	.44
	8(Terry)	1	.98	1.00
	9(Vickie)	1	27.16***	4.82**
	10(Bob)	1	28.48***	1.16
3	1(Steven)	2	71.74***	2.36
	2(Ruth Ann)	2	70.26***	7.47**

TABLE 11 CONTINUED

Subtest Number	Test Item	Degrees Freedom	Chi-Square to Subtest Median	Chi-Square to Total Score Median
4	1(A)	1	99.74***	14.89***
	2(B)	1	49.76***	12.12***
	3(C)	1	21.67***	6.94***
	4(D)	1	94.90***	12.74***
	5(E)	1	22.89***	10.49***
5	1(John)	2	33.56***	15.03***
	2(Mary)	2	23.42***	4.96**
	3(Karen)	2	15.38***	.37
	4(Eddie)	2	17.57***	2.68
	5(Pam)	2	53.70***	2.72
7	1(thine)	1	70.41***	17.45***
	2(fig)	1	63.19***	13.48***
	3(crush)	1	13.21***	10.30***
	4(rock)	1	38.45***	11.09***
	5(zeal)	1	39.91***	10.30***
	6(code)	1	45.16***	6.34**
	7(croak)	1	52.30***	8.93***
	8(plxy)	1	35.00***	3.38*
8	1(sidewalk)	1	8.80***	6.61**
	2(close)	1	33.33***	9.18***
	3(winner)	1	2.57	3.7*
	4(calling)	1	14.15***	1.45
	5(brighten)	1	20.66***	2.95*
	6(bacon)	1	39.06***	6.36**
	7(cabin)	1	36.43***	12.35***
	8(unafraid)	1	9.44***	.04
	9(reheat)	1	24.31***	2.45
	10(distant)	1	24.00***	2.57
	11(nobleman)	1	41.69***	12.38***
	12(breezy)	1	11.30***	5.11**
9	1(1&5)	1	59.15***	20.72***
	2(2-6-10, 3 pts)	1	28.38***	6.02**
	3(2-6-10, 2 pts)	1	20.97***	3.77*
	4(2-6-10 2 or 3 pts)	1	71.69***	14.22***

TABLE 11 CONTINUED

Subtest Number	Test Item	Degrees Freedom	Chi-Square to Subtest Median	Chi-Square to Total Score Median
	5(3&8)	1	93.21***	15.65***
	6(4-9)	1	79.26***	22.75***
	7(7)	1	82.50***	16.93***
10	1(Beginning Sounds)	1	20.03***	.00
	2 "	1	14.99***	2.77*
	3(Context)	1	27.36***	3.33*
	4 "	1	28.81***	9.33***
	5(Middle Sounds)	1	17.23***	4.90**
	6(Double Sounds)	1	54.74***	2.27
	7(Structural Analysis)	1	17.95***	3.84**
	8(Final e)	1	50.96***	.36
	9(Structural Analysis)	1	18.88***	9.28***

*Significant at the .10 level
**Significant at the .05 level
***Significant at the .01 level

Of the 70 chi-square values for each item's relationship to the subtest score, only six fail to reach at least the five percent level of confidence. Of these six, three are significant against the total score and one is essential to problem two. That leaves only two of the 70 that should be discarded. The advantage gained by discarding the two items would not be worth the loss of the norms.

It is interesting to note the discriminatory power of the subtests as a whole against the total score. Subtest four, where the teacher ranked five printed excerpts for difficulty, appeared to be the most powerful, with all five of its items significant beyond the one percent level. Subtests seven (phonic rules) and nine (concept formation on word perception errors) were of about equal power. Subtest two, forming reading groups, appeared to be the weakest.

To be a homogeneous test each subtest should correlate significantly with the total score. This was crucial for subtests one, two, three, and five which did not hold up well on the gain-with-experience predictions. The correlations are shown in Table 12.

TABLE 12. CORRELATION COEFFICIENTS OF EACH SUBTEST WITH THE TOTAL SCORE FOR ALL SAMPLES.

Subtest	Under-graduates	Student Teachers	Southern Indiana	Chicago	Combined Teachers
1	.41	.38	.38	.44	.40
2	.24*	.55	.19*	.59	.41
3	.47	.30	.27	.40	.33
4	.50	.53	.47	.54	.49
5	.31	.54	.51	.38	.43
7	.40	.62	.57	.58	.56
8	.56	.54	.38	.58	.48
9	.62	.49	.49	.71	.57
10	.52	.58	.42	.57	.55

*Significant at the .05 level
Unmarked correlation coefficient significant beyond .01 level.

One very interesting aspect of Table 12 is that in the Chicago sample, correlations ran much higher than in the southern Indiana sample. This is markedly apparent in problem two. A difference, though not as apparent, occurs between the student teachers and the undergraduates. To explain this, one much look first at the variances of the total score for the groups as shown in Table 13.

TABLE 13. TOTAL SCORE VARIANCES FOR ALL SAMPLES.

Sample	Variance	F Ratio	Significance
Chicago	117.01	2.02	$p < .01$
Southern Indiana	58.00		
Practice Teachers	98.19	1.52	$p < .05$
Undergraduates	64.56		

One possible interpretation of the results in Table 13 is that the administration and selective migration in the southern Indiana system have worked together to produce a more homogeneous group of teachers. It is possible that many skillful teachers left of their own choice and many incompentent ones left by the administration's choice. Those remaining probably do not constitute a representative sample. Smaller variances mean poorer item and subtest analyses as well as lower predictive value for teacher variables such as intelligence, attitudes and problem solving skill.

Perhaps one explanation for the increased variance of the student teachers is that the skillful ones increase their skills in solving reading problems at a much faster rate during the twelve week training period than did the less skillful.

Factor analysis. The correlation matrix from which the factor analysis was computed is shown in Table 14.

TABLE 14. CORRELATION MATRIX OF COMBINED TEACHER SAMPLE. N = 176

Subtest	1	2	3	4	5	7	8	9	10	Total
1		.10	.06	.04	.11	.18	.18	.01	.13	.40
2			.07	.13	.17	.15	.07	.14	.28	.41
3				.23	.02	.22	.00	.08	.25	.33
4					.27	.16	.09	.18	.10	.49
5						.10	.01	.00	.13	.43
7							.25	.07	.33	.56
8								.21	.13	.48
9									.18	.57
10										.55

Table 15 shows the eigenvalues which were derived, as the first step in the analysis, by use of the principal axis method.

TABLE 15. EIGENVALUES FOR EACH SUBTEST AND TOTAL.

Subtest	Eigenvalues
1	0.7922
2	0.9663
3	1.0715
4	0.5517
5	1.1625
7	0.5997
8	0.6990
9	1.0740
10	3.0820
TOTAL	0.0012

As suggested by H.F. Kaiser, only those four eigenvalues (and their associated principal components) which exceeded one were selected for rotation Table 16 shows the results of the Illiac varimax rotation program.

Factor loadings which exceeded .302--the .01 level of confidence for 67 degrees of freedom--were removed and placed in Table 17 with a subjective description of the process which the author believed was involved in the solution of each subtest. The confidence level was based on the 67 degrees of freedom of the Chicago sample rather than the 174 of the combined teacher group because of the heterogeneity of the combined sample.

TABLE 16. SUBTEST FACTOR LOADINGS, FACTOR VARIANCES, AND COMMUNALITIES FROM THE VARIMAX ROTATION.

Subtests	Factor Loading 1	2	3	4	Communalities
1	-.038	.136	.747**	.096	.5864
2	.089	.413**	.177	.231	.2731
3	-.034	.058	-.222	.798**	.6910
4	.270	.647**	-.237	.195	.5853
5	-.121	.849**	.167	-.083	.7696
7	.120	.075	.427**	.582**	.5414
8	.631**	-.073	.508**	-.005	.6624
9	.852**	.112	-.154	.109	.7737
10	.122	.145	.230	.659**	.5231
Sum of factor Variances	1.252	1.375	1.239	1.530	

**Significant beyond .01 level

TABLE 17. DESCRIPTION OF THE SOLUTION PROCESS INVOLVED IN THE SUBTESTS WHICH LOAD ON EACH FACTOR.

Factor	Subtest	Loading	Solution Process
A	9	.852	Teacher forms concepts then groups the word perception errors.
	8	.631	Teacher chooses the syllabication rule or concept to which each error belongs.
B	5	.849	Teacher estimates child's oral reading level then matches him to suitable reading material.
	4	.647	Teacher estimates difficulty level of printed reading excerpts then ranks them.
	2	.413	Teacher estimates child's oral reading level then ranks him with other readers.
C	1	.747	Teacher compares a child's reading to her own criterion of acceptability in selecting the reading book of suitable difficulty. The significance of errors and the stability of the teacher's criterion are factors.
	8	.508	The teacher writes the rule which will correct the syllabication errors.
	7	.427	Teacher recognizes phonic error and supplies the rule to correct it. In some cases, the most significant error must be selected.
D	3	.798	Teacher judges amount of difference between two identical passages read one after the other.
	10	.659	Teacher must recognize similarities and differences in intent between printed reading exercises.
	7	.582	Teacher must recognize differences between a reader's phonetic attack and the correct pronounciation of words, then supply the rule to correct the error.

A matrix of associations between these four factors and many other variables should be prepared before the factors can be named with any degree of security. At this time, with awareness of the risks involved, some tentative guesses are made. Factor A may reflect a teacher's ability to form concepts and make sense out of a child's errors. Factor B may reflect a teacher's ability to recognize the level at which a child is reading. Factor C may reflect the teacher's skill in providing the proper learning experience, i.e., the correct reading book or word attack rule. The skill which is reflected by factor D is less evident than that of A, B, or C. In general, factor D appears to be a skill in perceiving varying degrees of similarity among readers and reading exercises. Further elaboration will have to come from research in which the four factors are variables.

Relationship between the reading-problems test and other variables. If the reading problems account for some of the variance of teacher performance in the classroom, it is of interest to know what characteristics of the teacher are associated with test performance. This is primarily important for the selection of teacher candidates. Table 18 shows the relationships among the reading problems and four other variables for student teachers.

TABLE 18. CORRELATION MATRIX OF THE TOTAL SCORE OF THE READING-PROBLEMS TEST, ACE, CO-OP C-2, MTAI, AND GRADE-POINT AVERAGE ON 45 INDIANA UNIVERSITY STUDENT TEACHERS.

Variable	ACE	CO-OP	MTAI	Grade-Point
Reading-Problems Test	.56**	.55**	.08	.01
ACE		.78**	.01	.37**
Co-op C-2			.12	.45**
MTAI				.03

** Significant beyond .01 level

From Table 18, the two student teacher variables which are associated with performance on the reading problems are intelligence (ACE) and reading comprehension (Co-op C-2). The substantial correlations are due in part to the variability of the practice teacher sample; the sample standard deviation on the ACE is 24 points, which is only two points less than that of the standardization population. (Probably because of the homogeneity of teacher samples, researchers find only low correlations between intelligence and success.) The high correlation between the ACE and the Co-op C-2, suggests a good deal of communality and probably accounts for their nearly identical relationship with the reading-problems test. Nevertheless, whatever they are measuring,--reading skill and/or intelligence- -their

correlation with the reading-problems test is tenable. The correlations could be used as two more pieces of evidence for the validity of the reading-problems test if two assumptions are granted. First, as a group, teachers who are skillful in solving classroom problems in reading instruction are more intelligent than poor problem solvers. Second, some of the reading instructional skills used by teachers may result, in part, from transfer from the teachers' own reading abilities.

The lack of relationship with the MTAI is not surprising in a student teacher sample. There is no reason why these persons with little experience, who have attitudes which should lead to rapport with pupils, should necessarily be skillful in solving reading problems. It would be more reasonable to expect a correlation between MTAI and the reading-problems test in teachers with several years of experience. Those experienced teachers with permissive attitudes and rapport-establishing skills are likely to be the ones who sit down with their children and give specific and individual help, thereby developing skills measured by the reading-problems test. If a study were done longitudinally, it would be worthwhile to investigate the gains made on the reading-problems test by teachers who score high and low on the MTAI.

The zero correlation with grade-point average, if it is consistent on replication, suggests that the validity of college marks for the selection of skillful teachers of reading is questionable.

SUMMARY

The object of this study was to construct and begin to validate a reliable normative test of a sample of teacher skills involved in reading instruction in grades 2 through 5.

Preliminary validation of the reading problems test was done in two ways. First, the hypothesis was made that teachers become more skillful in solving teaching problems as they gained experience. A valid test of teacher skills in reading should therefore reflect this gain in skill with experience. Samples of teachers, student teachers, and education undergraduates were tested and cross-sectional comparisons were made on each of the nine subtests as well as on the total score. Second correlations were computed between the reading-problems test and selected independent variables, particularly intelligence and reading comprehension. Table 19 presents a summary of the results.

TABLE 19. SUMMARY OF RESULTS OF THE VALIDATION, RELIABILITY, AND INTERNAL CONSISTENCY STUDIES FOR THE TOTAL SCORE AND SUB-TESTS.

Prediction	Total Score	Sub-test 1	2	3	4	5	7	8	9	10
Southern Indiana Teachers Outscore Undergraduates.	yes	no	yes	no	yes	no	yes	yes	yes	yes
Chicago Teachers Outscore Undergraduates. (Cross-Validation)	yes	no	no	no	yes	no	yes	yes	yes	yes
Student Teachers Outscore Undergraduates.	yes	no	no	no	no	no	yes	yes	yes	yes
Combined Southern Indiana and Chicago Teachers Outscore Student Teachers.	yes	no	no	no	no	no	yes	yes	no	no
Teachers 16-30 Years Experience Outscore Teachers 1-15 Years Experience in Chicago Sample.	yes	---- not compared ----------								
Split-half Reliability Above .85 Chicago Teacher Sample.	yes	------not determined ---------								
Each Sub-test will Correlate Significantly with Total Score Across All Samples.	---	yes	yes	yes	yes	yes	yes	yes	yes	yes

Note: "yes" means that the difference was significant at least at the .01 level.

BIBLIOGRAPHY

1. Bond, Guy L., and Tinker, Miles A., Reading Difficulties, Their Diagnosis and Correction, Appleton-Century-Crafts, Inc., New York, 1957.

2. Cook, Walter W.; Leeds, Carroll H.; and Callis, Robert, Manual, Minnesota Teacher Attitude Inventory, The Psychological Corporation, New York, New York, undated.

3. Davis, Charlotte C., Cooperative English Test, Test C-2, Reading Comprehension, Higher Level, Form Z, Educational Testing Service, Cooperative Test Division, Princeton, New Jersey, 1953.

4. Gates, Arthur I., "Teaching Reading" in the series What Research Says To The Teacher, Department of Classroom Teachers, American Educational Research Association of the National Education Association, Washington 6, D.C., June, 1953.

5. Gilmore, John V., The Gilmore Oral Reading Test, World Book Company, New York, 1951.

6. Gray, William S., On Their Own in Reading, Scott, Foresman and Company, Chicago, 1948.

7. Greene, Harry A., and Gray, William S., "The Measurement of Understanding in the Language Arts," in The Measurement of Understanding, pp. 175-212, Forty-fifth Yearbook, National Society for the Study of Education, Public School Publishing Co., Bloomington, Illinois, 1946.

8. Russell, David H., and Karp, Etta E., Reading Aids Through the Grades, Bureau of Publications, Teachers College, Columbia University, 1950.

9. Thurstone, L.L., and Thurstone, Thelma G., American Council on Education Psychological Examination for College Freshman, Educational Testing Service, Cooperative Test Division, Princeton, New Jersey, 1952.

10. Turner, Richard L., and Fattu, Nicholas A., Skill in Teaching, A Reappraisal of the Concepts and Strategies in Teacher Effectiveness Research, Bulletin of the School of Education, Indiana University 36:3, May, 1960.

11. Turner, Richard L., and Fattu, Nicholas A., Problem Solving Proficiency Among Elementary School Teachers I. The Development of Criteria, Monograph of the Institute of Educational Research, Indiana University, Bloomington, May, 1960.